# Hybrid Multi-Objective Active Appearance Model for Gamer's Facial Features Detection

Abdul SATTAR, Renaud SEGUIER

SUPÉLEC / IETR
Avenue de la Boulaie, 35576 Cesson Sévigné, France
abdul.sattar@supelec.fr, renaud.seguier@supelec.fr

**Résumé** – Nous proposons dans cet article un algorithme d'analyse facial robuste multi-caméras pour les jeux vidéo, capable d'extraire des caractéristiques faciale lorsque le visage produit de larges mouvements latéraux. La gestuelle des joueurs lors des manuvres tactiques produites dans les mondes virtuels rend l'analyse des visages impossible avec une seule caméra. Notre système d'acquisition multi-caméras permet de résoudre ce problème mais en soulève un autre : celui de l'analyse d'un même visage par plusieurs sources vidéo synchrones. Nous proposons une telle analyse en nous appuyant sur des modèles actifs d'apparence 2.5D pour lesquels nous proposons une optimisation hybride (algorithmes génétiques et descente de gradient) dans le contexte multicritères de Pareto. Cette proposition est évaluée sur des séquences multi-vues de visages réels et synthétiques, des tests comparatifs avec une approche non-hybride illustre son intérêt.

**Abstract** – In this article we propose a robust facial analysis of multiple camera system for cyber games which is capable of extracting the facial features of a face making large lateral movements. Tactical maneuvers of the gamer make single camera acquisition system unsuitable to analyse and track the face due to his large lateral movements. Although our proposition of double camera acquisition system resolved this problem, but the facial data obtained from both cameras produces optimization problems for face search algorithms. For an improved facial analysis system, we propose to acquire the facial images from two cameras and analyse them by Pareto based hybrid multi-objective face search optimization for 2.5D Active Appearance Model (HMOAAM). Proposed algorithm is applied on number of multi-view real and synthetic facial images and its results are compared with a non hybrid system. Results obtained validate our proposition.

## 1 Introduction

Over the last decade, cyber games have became more and more interactive. Gamers need to get involved in the game to an extent to visualize opponent's face and interact with him virtually. To track and analyse gamer's face efficiently and to ensure his interaction, system needs to overcome the bottlenecks in facial analysis. Facial analysis deals with the face alignment, pose, features, gestures and emotions extractions. Excitements caused by the tactical maneuvers in a game, compel the gamers to move around in various directions. These maneuvers produce large lateral movements of a face, which makes it difficult for a facial analysis system to track and analyse the face.

In single-view system, face cannot be analysed when it occludes itself during its lateral motion. Such as in a profile view only half of the face is visible. To overcome this dilemma we exploit facial information from another camera and associate it with the one unable to localize at the first place. This association helps the search methodology to reduce the possibility of divergence. Moreover better outcomes of one camera can escort the other. In multi-view system, optimization of more than one error is to be performed between a model and query images from each camera. Searching for an optimum solution of a single task employing two or more distinct errors requires multi-objective optimizations (MOO). In this paper we propose Pareto based NSGA-II [3] hybridized with gradient descent (GD) method.

## 2 Related Work

Active Appearance Model (AAM) presented by [1], is one of the well known efficient method in feature extraction and alignment of a face. Researchers have developed various techniques for the improvement of face analysis by AAM. Cootes et al. and Ting et al. tackled the problem of large lateral movements of a face by using 3 AAM models in [2, 11], one dedicated to the frontal view and two for the profile views. Use of more than one model of AAM has various drawbacks, which includes high memory usage for the storage of shapes and textures of all the models, extensive computations requirement to determine the model required for query images. Hu et al. proposed a robust algorithm of fitting a 2D+3D AAM to multiple images acquired at the same instance in [5]. Their fitting methodology, instead of decomposing into three independent optimizations from three cameras, adds all the errors. Kim et al. proposed another algorithm of face tracking by Stereo AAM (STAAM) fitting in [6], which is an extension of the above fitting of 2D+3D AAM to multiple images. They used gradient descent (GD) algorithm alone as a fitting method, which eventually requires to pre-compute Jacobian and Hessian matrices. Moreover lack of exploration capability of the method makes GD very sensitive to initialization.

As far as hybrid optimization is concerned, Simplex and Genetic Algorithm (GA) were combine in [4], whereas evolutionary algorithm (EA) with gradient based algorithms were combined in [9] for single objective optimizations. For MOO methods [13, 8] combined GA with Simplex and Simulated Annealing respectively and [12, 7] combined EA with gradient based algorithms. Hybridization of two algorithms has shown promising results. In the next section we propose an efficient and robust algorithm for facial features extraction by 2.5D AAM. Our proposition not only eliminates the steps of precomputation but is also insensitive to initialization.

# 3 Hybrid Multi-Objective Active Appearance Model (HMOAAM)

As explained in section 2 some researcher have tried to analyse the face by emphasizing on the model generation and their search methodologies, while others emphasized on increasing the facial information by using multiple cameras. However instead of treating multiple camera AAM face search as a multi-objective optimization they performed addition of the errors from each camera eventually making it a single objective optimization task. In this section we present our proposed hybrid multi-objective double camera facial analysis system of HMOAAM. In addition to multi-objective nature, HMOAAM is also capable of analysing the unseen and unknown faces.

## 3.1 2.5D Active Appearance Model

We have used the same 2.5D AAM of [10] which is constructed by i) 2D landmarks of frontal and profile views of the facial images combined to make 3D shape and ii) 2D textures of only frontal view of the facial images mapped on the mean 3D shape. Principal Component Analysis (PCA) is performed on these shapes and textures to obtain their respective parameters $b_s$ and $b_g$ with 95% of the variation stored in these parameters. Both of these parameters are concatenated and a final PCA is performed to obtain the appearance parameters $C$. This 2.5D AAM can be translated as well as rotated with the help of pose vector given as

$$P = [\theta_{pitch}, \theta_{yaw}, \theta_{roll}, t_x, t_y, Scale]^T$$

## 3.2 HMOAAM Fitting

The objective of HMOAAM fitting is to minimize error between segmented images from each camera and the model instance by varying $C$ and $P$ parameters. In the first phase of HMOAAM parameters $C$ and $P$ are optimized by NSGA-II which is a multi-objective version of GA. $C$ and $P$ parameters are concatenated to form population of various chromosomes. Each chromosome contains similar $C$ parameters for both facial query images, whereas in $P$ parameters $\theta_{yaw}$ has an offset equal to the angular distance between two cameras. Hence each chromosome represents two AAM model as $M_1$ and $M_2$. In segmentation this deformed, rotated and translated shape models are placed on both query images from each camera to warp the face to mean frontal shape. Two pixel errors (fitness) are calculated between the warped query images from each camera and their respective models ($M_1$ and $M_2$) represented by a single chromosome. Tournament selection is applied to select parents from the population to undergo reproduction. Two point crossover and Gaussian mutation is implemented to reproduce next generation of the chromosomes. During mutation of each parameter we also calculate the gradient of the respective parameter of the chromosome (explained later). As each chromosome contains a solution to two query images along with their pixel errors therefore, for the reproduction and selection, non-dominating scenario is to be implemented to obtain the desired Pareto optimal solution. The basic idea is to find the set of solutions in the population that are Pareto non-dominated by the rest of the population. These solutions are assigned the highest rank and are removed from further assignment of the ranks. Similarly remaining population undergoes the same process of ranking until the population is suitably ranked in the form of Pareto fronts.

At the end of this first phase of HMOAAM we are able to form Pareto fronts of the solutions from both cameras. In other words we have now explored and exploited facial search space. GA based algorithms are known to find the region of the global optimum values instead of precise global optimum because of their exploration quality better than their exploitation. To improve the exploitation of our system we have hybridized NSGA-II with GD which is a highly deterministic algorithm and have high tendency to fall in local minimum. Applying GD alone will not serve our purpose since facial search space contains large number of local minima.

In the second phase of HMOAAM we have implemented GD on the solutions obtained at the end of first phase. Chromosomes of first Pareto rank are sorted with respect to their euclidean distance from center of gravity of the chromosome population. Three best chromosomes are selected for the GD to find local minima. In order to apply GD we need Jacobian matrices, which are the partial differentials of the error function with respect to each parameter. Simple integration of GD in NSGA-II would have increased the number of error evaluations, but we have proposed gradient operator which functions in conjunction with mutation operator. Thus gradient operator uses the error evaluation of mutation operator and do not put an extra burden on the system. During mutation of a chromosome only one parameter is changed for next generation. Error evaluation of this variation is retrieved by gradient operator as a partial differential of a function to build $J_C$ (Jacobian matrix of $C$ parameters) and $J_P$ (Jacobian matrix of $P$ parameters). These Jacobian matrices direct their respective parameter to optimum solution by pointing to the variation of each parameter. Therefore changes in $C$ and $P$ parameters are calculated by multiplying the transpose of respective Jacobian matrix with the residual images obtained by the solutions of the first phase. The equations of calculating $\Delta C$ and $\Delta P$ are given as

$$\Delta C = -\eta J_C^T e_x \tag{1}$$

$$\Delta P = -\eta J_P^T e_x \tag{2}$$

where $\eta$ is the step size to control the variations of the parameters in the direction of the gradient and $e_x$ is the residual image. Equations 1 and 2 are used to calculate $\Delta C$ and $\Delta P$ for four values of step size i.e. 0.25, 0.5, 0.75 and 1.0. At the end of second phase the best $C$ and $P$ parameters given by GD is selected and ground truth is calculated for the comparison of results. The gradient operator does not introduce significant time delays since it retrieves information from another operator. Moreover Jacobian matrices are not that huge to store while evaluating solution.

# 4 Experimental Results

For the experimental results we have developed a real face database, from a facial image acquisition system composed of two webcams installed at the extreme edges of a screen with the angular distance of 50°. Same scenario is implemented in a software MAYA to obtain database of synthetic facial[1] stereo images. In learning phase 2.5D AAM model is build using the frontal and profile view facial images of a well known database of M2VTS[2].

For the first test database we have taken 266 facial images (all the images from left semi profile to right semi profile) of 7 individuals of real face database. Similarly for the second test database we have taken 4800 facial images (all the images from left semi profile to right semi profile) of 60 individuals of synthetic face database.

The results from both real and synthetic test databases are shown in figures 1 and 2 respectively. In figure 1, with a ground truth error less than 15% of the distance between the eyes, HMOAAM is able to extract facial features of 61% of the total real facial images, compared to 48% by MOAAM. Similarly in figure 2, with a ground truth error less than 15% of the distance between the eyes, HMOAAM is able to extract facial features of 60% of the total synthetc facial images, compared to 45% by MOAAM. Both figures depict that our system of HMOAAM fitting is lot better than MOAAM fitting.

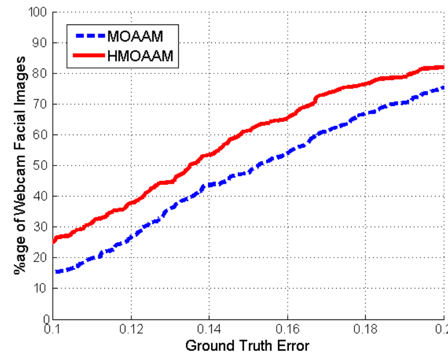


FIG. 1: HMOAAM vs. MOAAM for real faces.

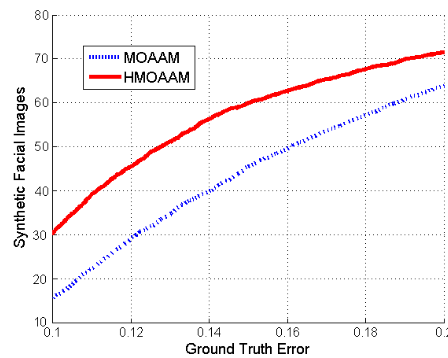


FIG. 2: HMOAAM vs. MOAAM for synthetic faces.

Comparison of localization of facial features (eyes, nose and mouth) for both real and synthetic facial images are shown in figures 3(a) and 3(b) respectively. It can be seen from both figures that facial features are well localized by HMOAAM than MOAAM.

As far as time consumption is concerned, HMOAAM requires 3000 warps to extract features of an oriented face. Single warp in an iteration is equal to 90% of the time consumed by an iteration i.e. 0.03 msec in Pentium IV 3.2GHz. Thus for a complete facial analysis of a face HMOAAM requires 100 msec, which means it can successfully analyze 10 frames in one second. Method of MOAAM is also restricted to complete its facial search with in this time period in order to compare its robustness and efficiency with HMOAAM.

[1]Synthetic faces were made in a software named as "Facial Studio". All of them were imported in MAYA for rendering the synthetic facial images.

[2]M2VTS: Multi Modal Verification for Teleservices and Security applications. http://www.tele.ucl.ac.be/M2VTS/


(a) Real Faces (b) Synthetic Faces

Fig. 3: Comparison of feature localization by HMOAAM and MOAAM. (a) Real faces. (b) Synthetic faces.

It is important to point out that the above mentioned computational time is required to align the face without any prior knowledge of the facial pose, however in a face tracking mode the required time is reduced enormously by employing pose parameters of the previous frames, thus making it a true real time application.

# 5 Conclusion

Large lateral movements of a gamer in front of the display make it impossible to track and analyse his face with single camera. In this article we have presented an efficient technique to overcome this bottleneck. We have given the solution using double camera acquisition system. Face analysis in this system is implemented by a robust and efficient algorithm of Pareto based NSGA-II hybridized with GD for 2.5D AAM. With the help of the results we have shown that our proposition of HMOAAM outperforms non hybrid technique of MOAAM. For the moment, this approach is limited to be used in an interactive system for the gamers, but it would be interesting to extend it for larger events, like conferences and meetings, with multiple cameras installed on different corners of the room and displayed on video projectors.

# References

[1] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *European Conference on Computer Vision*, volume 2, pages 484–498, 1998.

[2] T. Cootes and C. Taylor. Statistical models of appearance for computer vision, 2004. Technical report from Imaging Science and Biomedical Engineering, University of Manchester.

[3] K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. *Parallel Problem Solving from Nature VI Conference*, pages 849–858, 2000.

[4] N. Durand and J. m. Alliot. A combined nelder-mead simplex and genetic algorithm. *The Genetic and Evolutionary Computation Conference*, pages 1–7, 1999.

[5] C. Hu, J. Xiao, I. Matthews, S. Baker, J. Cohn, and T. Kanade. Fitting a single active appearance model simultaneously to multiple images. *British Machine Vision Conference*, 2004.

[6] D. Kim and J. Sung. A real-time face tracking using the stereo active appearance model. In *Proc. IEEE International Conference on Image Processing*, pages 2833–2836, 2006.

[7] M. Lahanas, D. Baltas, and N. Zamboglou. A hybrid evolutionary algorithm for multi-objective anatomy-based dose optimization in high-dose-rate brachytherapy. *Physics in Medicine and Biology*, 48:399–415, 2003.

[8] G. Li, G. Song, Y. Wu, J. Zhang, and Q. Wang. A multi-objective hybrid genetic based optimization for external beam radiation. *27th Annual International Conference of the Engineering in Medicine and Biology Society*, pages 7734 – 7737, 2006.

[9] A. Manfred, F. Sonke, R. Karl, and S. H. Siegfried. Hybrid optimization for 3d landmark extraction: Genetic algorithms and conjugate gradient method. *Bildverarbeitung fur die Medizin*, pages 314–317, 2002.

[10] A. Sattar, Y. Aidarous, and R. Seguier. Gagm-aam: A genetic optimization with gaussian mixtures for active appearance models. In *Proc. 15th IEEE International Conference on Image Processing ICIP 2008*, pages 3220–3223, 2008.

[11] S. Ting, C. L. Brian, and C. Shaokang. Face recognition robust to head pose from one sample image. *18th International Conference on Pattern Recognition (ICPR'06)*, pages 515–518, 2006.

[12] E. Wanner, F. Guimaraes, R. Takahashi, D. Lowther, and J. Ramirez. Multiobjective memetic algorithms with quadratic approximation-based local search for expensive optimization in electromagnetics. *IEEE Transactions on Magnetics*, 44:1126 – 1129, 2008.

[13] S. Zapotecas Martinez and C. A. Coello Coello. A proposal to hybridize multi-objective evolutionary algorithms with non-gradient mathematical programming techniques. *10th international conference on Parallel Problem Solving from Nature*, pages 837–846, 2008.